

# SEGMENTATION AND LABELLING OF INTRA-OPERATIVE LAPAROSCOPIC IMAGES USING STRUCTURE FROM POINT CLOUD

*Nazim Haouchine and Stephane Cotin*

Inria, Mimesis project-team, Strasbourg, France

## ABSTRACT

We present in this paper an automatic method for segmenting and labelling the liver and its surrounding tissues in intra-operative laparoscopic images. The goal is to distinguish between the different structures that compose a intra-operative hepatic surgery scene to improve common registration tasks. Our segmentation method considers the scene as a 3D structured point cloud to exploit powerful informations such as curvature and normals, in addition to visual cues that permit to efficiently classify the scene. Experiments performed on challenging human hepatic surgery confirm that accurate segmentation and labelling are possible using our approach.

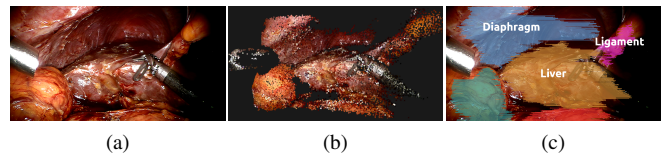
**Index Terms**— Intra-operative image segmentation, minimally invasive surgery, scene understanding.

## 1. INTRODUCTION

The last decades have seen the emergence of Minimally Invasive Surgery (MIS). In this technique, the surgeon manipulates instruments inserted in the the abdominal cavity by observing a display showing a video stream captured by an endoscopic camera inserted through the navel. The main advantage of the method is the reduction of pain, time recovery and risks of infection. However, although the MIS procedure is considered as a well-established technique, it remains quite complex from the surgical standpoint mainly due to the limited visual feedback and indirect manipulation. The presence of a camera has naturally led the research community to investigate the use of Augmented Reality (AR) to guide the surgeon during the procedure. Indeed, organ structures computed from pre-operative scans can be superimposed onto the intra-operative images to facilitate the orientation and navigation during the surgery.

Several approaches have been proposed to permit the use of AR during surgery. Most of the time they rely on a registration step which is often performed between a pre-operative model (usually a triangular 3D mesh) and points of interest (2D or 3D features) extracted from laparoscopic images [1], [2], [3]. Although these methods provide good results, the registration process suffers from a lack of image information and can hardly achieve the robustness needed by surgical applications. Recently, researchers have suggested the integra-

tion of contours and anatomical landmarks as additional constraints for the registration phase [2], [3]. These landmarks represent powerful boundary conditions that can dramatically decrease the registration error and resolve ambiguities where no image information is provided.



**Fig. 1:** Segmentation and labelling of MIS hepatic scene: (a) input laparoscopic images, (b) 3D scene reconstruction, (c) segmentation and labelling of anatomical structures

## 2. RELATED WORKS

Detecting and extracting regions of interest in laparoscopic images has first been dedicated to the localisation of surgical instruments. Being an important element of advanced robotic and computer assisted interventions, several methods have been proposed to address this problem [4], [5] with significant results. However, when dealing with real organs and tissues, the classification process become more challenging due to the substantial intra-patient organ texture disparity.

Early works in that context were based on texture and morphological similarity, using iterative mergin process based on grey-level similarity [6] or watershed segmentation in HSI color space [7]. These methods do not guarantee success and convergence since they are highly dependent on the texture.

In order to counteract the sensitivity to color and texture, learned models have been recently introduced. A cascaded patient-specific segmentation method is proposed in [8]. This method exploits both color and texture encoded using Support Vector Machines to segment the uterus and Gaussian Mixture Model to segment the tools. The results show that segmenting the uterine image can improve 3D reconstruction of the uterus and its registration with pre-operative data. In a similar context and towards the establishment of automatic

AR for uterine surgery, Prokopets *et al.* [9] proposed to detect the uterus and fallopian tube junction in laparoscopic images. This method is based on a learned contextual approach that aims to predict the positions and orientations of the junctions relative to the uterus using a statistical model (obtained through a training phase). These approaches give good results using only single view images. However, the training phase may be computationally expensive and requires a large database, which makes the solution not easily reproducible with other organs.

Using pre-operative data as well as endoscopic visual cues Nosrati *et al.* [10] proposed a closed form solution to segment multiple tissues of partial nephrectomy images. This method makes use of multi-view endoscopic videos and the 3D segmentation of the organ in pre-operative scans to formulate the problem as a 3D pose search. The presented results as well as the real-time achievement make this solution suitable for surgical AR. Nevertheless, the use of a deformable model generated from a catalog of realistic 3D deformed shapes does not represent the patient-specific tissue deformations, especially when dealing with elastic organs such as the liver.

We propose in this work to automatically classify the structures that compose an intra-operative hepatic surgery scene. We use geometrical discontinuities to find the boundaries between each region and structural cues to label them. In contrast to the state-of-the-art, our method does not rely on learned or pre-operative models and is compatible with multiple organ scenes with intra-patient disparity.

### 3. METHODS

#### 3.1. Intra-operative 3D Scene Reconstruction

The reconstruction of a three-dimensional intra-operative scene can be performed in different ways depending on the image modality [11]. In this work, we make no assumptions about specific acquisition technologies available in the operating room (such as intra-operative MR or CT scanner, or depth camera), but we rely only on images provided by the stereoscopic camera (pre-calibrated).

Let us assume the stereoscopic pair of images  $\mathcal{I}_l$  and  $\mathcal{I}_r$ . We extract from this stereo pair points of interest (features) that are sufficiently reliable for 3D reconstruction using a feature detector. This permits to obtain two sets of features  $F_l = (f_{l_1}, \dots, f_{l_k})$  and  $F_r = (f_{r_1}, \dots, f_{r_m})$  where it is necessary to estimate for each feature  $f_i = (u_i, v_i)$  the 3D point  $p_i = (U_i, V_i, W_i)$ . This is done by establishing a correspondence between image points  $f_l \longleftrightarrow f_r$ , using a descriptor-based matching method with the appropriate threshold  $\phi$ . Once a correspondence is found, a sparse set of 3D points, denoted  $\mathbf{P} = \{p_i \in \mathbb{R}^3\}$ , is reconstructed using a triangulation algorithm [11].

The resulting point cloud  $\mathbf{P}$  gives a sparse representation of the intra-operative scene with possible noise measurement

coming mainly from calibration issue. We propose to use *Moving Least Square* (MLS) surface approximation introduced by [12] in order to build a dense surface from the sparse reconstruction obtained from stereoscopic images.

MLS is a non-parametric method to approximate a surface from a set of unstructured scattered point. This approach relies on the idea that any given point could implicitly defines a surface as the set of stationary points of a certain function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . This procedure is defined below.

Let consider the previous point cloud  $\mathbf{P}$ , and a point  $r$  in a neighborhood of  $\mathbf{P}$ . The energy of the plane with normal  $\vec{a}$  passing through the point  $x = r + t\vec{a}$ , where  $t$  is the distance from  $r$  to the plane, is defined to be:

$$\mathcal{E}_{MLS}(\vec{a}, t) = \sum_{p_i \in \mathbf{P}} (\langle \vec{a}, p_i \rangle - \langle \vec{a}, r + t\vec{a} \rangle)^2 \theta(p + t\vec{a}, p_i)$$

where the weighting function  $\theta$  is any monotonic function, usually a Gaussian. The energy  $\mathcal{E}_{MLS}$  measures the quality of the fit of the plane to  $\mathbf{P}$ , where  $p_i$  is weighted by its distance from  $x$ . The local minima of this energy function occurs at a discrete set of inputs  $(\vec{a}, t)$ , each corresponding to a point  $x$ . Of these,  $f(r)$  is defined to be the  $x$  nearest to  $r$ . The stationary points of this map  $f$  form the MLS surface  $\mathcal{S}$ .

#### 3.2. Segmentation using structures from Point Cloud

The obtained surface  $\mathcal{S}$  gives a 3D representation of the whole intra-operative scene. In order to separate the different component (regions) of the scene, we propose a classification method based on geometrical surface bending given by curvatures.

For each point on the surface  $\mathcal{S}$ , we can locally approximate the surface by its tangent plane, orthogonal to the normal vector  $n$ . Local bending of the surface is measured by curvatures. For every unit direction  $e_\theta$  in the tangent plane, the normal curvature  $k^n(\theta)$  is defined as the curvature of the curve that belongs to both the surface itself and the plane containing both  $n$  and  $e_\theta$ . The mean curvature  $k_H$  is defined as the average of the normal curvatures:

$$k_H = \frac{1}{2\pi} \int_0^{2\pi} k^n(\theta) d\theta \quad (1)$$

Using this curvature information, and knowing the bending characteristic of the intra-operative scene, the idea is to find strong bending discontinuities that represent the frontiers between the regions using a region growing scheme.

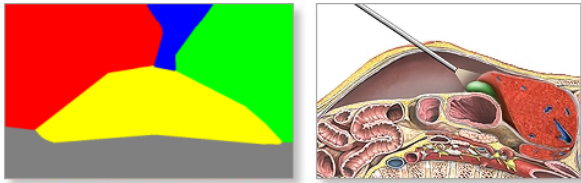
Let us denote  $\mathbf{V} = \{v_i \in \mathbb{R}^3\}$  the set of vertices of the surface  $\mathcal{S}$ . Region growing permits to classify the set of vertices  $\mathbf{V}$  starting from a set of seed points  $\mathbf{D} = \{d_i \in \mathbb{R}^3\}$  and growing through neighbors based on curvature criterion. The seed points are first chosen to be the points with the lowest curvatures  $k_H$ , representing the flat regions. From the seed point, the set of neighbours  $\mathbf{V}^* = \{v_i^* \in \mathbb{R}^3\}$  are examined so that the

angle between the seed point  $d_i$  and its neighbor  $v_j^*$  is below a certain threshold  $\vartheta_{th}$  following  $\cos^{-1}(|(n_{d_i}, n_{v_j^*})|) < \vartheta_{th}$ . In this case, the neighbor  $v_j^*$  is added to the current region. Once all the neighbors are tested, if the curvature of a neighbor is less than a curvature threshold  $k_{H_{th}}$ , this neighbor is added to the set of seed points  $\mathbf{D}$  while the seed point  $d_i$  is removed.

This process is first done using a strict matching threshold  $\varphi$  which reduces the density of the point cloud  $\mathbf{P}$ . This permits to avoid ambiguities to delimit the regions. Indeed, a dense point cloud  $\mathbf{P}$  will produce a smooth MLS surface  $\mathcal{S}$  with smooth angles, and therefore will cause the growing to build false regions. The growing is then repeated with more tolerant values of  $\varphi$  while the propagation is constrained with the initially detected regions. The process is iterated until all the vertices are labelled.

### 3.3. Labelling with laparoscopic scene cues

Just as there are many ways to parameterize an image using colors and texture of appearance, there are numerous ways to parameterize a 3D point cloud. When dealing with intra-operative scene, targeting hepatic resection, one will notice that the main anatomical structures are: the ligament, the liver, the diaphragm and the surrounding tissues (such as fat). Each of these structures can efficiently be labelled using structural cues. We propose several structural cues, chosen to be robust, intuitive and computationally cheap based on the previously reconstructed surface and the segmented regions. These cues are: distance to the camera  $\mathbf{t_D}$ , surface orientation  $\mathbf{t_O}$ , angular disparity  $\mathbf{t_A}$ , and anatomical connectivity  $\mathbf{t_C}$ .



**Fig. 3:** Intra-operative scene cues with in (left) the connectivity template for the cue  $\mathbf{t_C}$  and in (right) the camera insertion heading the lungs that permits to define  $\mathbf{t_D}$ ,  $\mathbf{t_O}$  and  $\mathbf{t_A}$ .

Combining these cues permits to label each of the aforementioned regions. However they are only valid under certain assumptions. First, we assume that the camera is inserted through the navel heading the lungs. Second, we suppose that the camera can observe at least one lobe. Finally, we assume that these structures are not significantly occluded by surgical instruments. The two first assumptions are most of time valid and the latter can be circumvented by taking advantage of the recent works in surgical tools detection [5]. Moreover, the cues are defined as follows.

**Distance to the camera  $\mathbf{t_D}$ .** This cue is very intuitive since it relates the distance in the world coordinates of the

anatomical structure to the camera position. The depth of each vertex of the surface is compared to the camera center's z coordinate in world coordinates. This cue is particularly useful to detect the ligament, which, in a classical intra-operative hepatic scene, is located close to the camera and the diaphragm, the deeper organ in the scene.

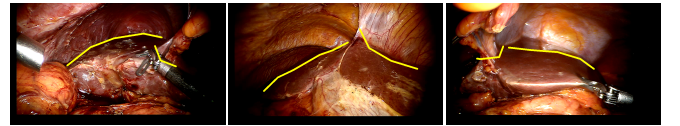
**Surface orientation  $\mathbf{t_O}$ .** The orientation of the surface normals gives a powerful insight knowing the location of the camera in the abdominal cavity (surrounded by the anatomical structures). This orientation cue helps to distinguish between the liver lower and upper boundaries.

**Angular disparity  $\mathbf{t_A}$ .** One can notice the "convex" bending of the organs (such as the liver, the kidney or the gallbladder) compared to the rest of the tissues. This geometry makes the exploitation of the angular disparity useful where a substantial angular disparity usually denotes the liver.

**Anatomical connectivity  $\mathbf{t_C}$ .** The connectivity between the regions is an important cue, since several structures share boundaries. Thus, the identification of one structure implies the identification of its neighbor. However, this cue is very sensitive and a false identification can lead to false implications and is used as a validation criterion rather than a decision criterion.

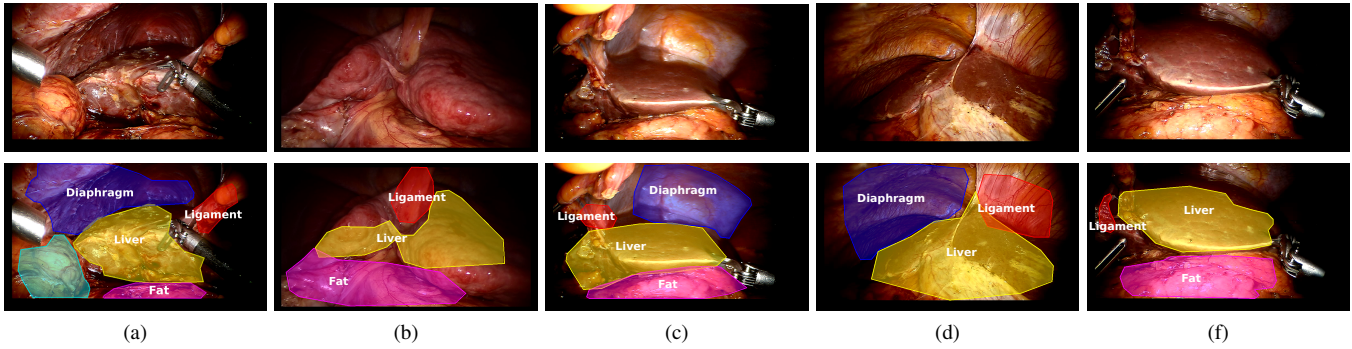
## 4. RESULTS

We conducted experiments on multiple laparoscopic image pairs from 3 different human patients showing the liver and its surrounding structures. Laparoscopic images of  $960 \times 540$  resolution were acquired using the DaVinci Robot (Intuitive Surgical) with a stereoscopic camera inserted through the navel heading the lungs. The images depict a classical hepatic resection scene where at least 3 anatomical structures are present. Our method was able to extract the liver contour with an average Dice similarity of 81.33 % while the labelling was able to detect it in all images as illustrated in Figure 2. This detection can bring powerful information when targeting 3D reconstruction of only one organ.



**Fig. 4:** Detection of the liver upper boundaries.

The experiments show that the boundary between the liver and the abdominal wall can be easily extracted when the diaphragm and the ligament are correctly detected (cf Figure 4). This separation represents a strong contour constraint towards an automatic pre-operative to intra-operative point cloud registration.



**Fig. 2:** Detection of liver, diaphragm, ligament and tissues on laparoscopic image pairs from haptic surgery on human patients.

The method however fails in certain cases to correctly separate the liver from its lower surrounding tissues such as fat (cf Figure 2 (d)). This can be explained by the non-presence of curvature discontinuities making the growing scheme fail to distinguish between the two regions.

## 5. CONCLUSION

We have presented an automatic method for segmenting and labelling intra-operative scene during liver surgery. This method is able to segment and label the liver, the diaphragm, the ligaments and surrounding tissues by considering the scene as a 3D structured point cloud to exploit curvatures and normals. Our approach can take as input sparse and noisy point clouds, thanks to a surface approximation stage, and unlike existing approaches, is independent of organs texture making it applicable to different organs. The experiments performed on real human hepatic surgery confirm that correct segmentation and labelling are possible using 3D structure information and appropriate visual cues. The major weakness of our method is its large number of parameters, where an inappropriate tuning may lead to incorrect results. Future work will investigate the automatic segmentation of instruments and the use of image information to resolve possible ambiguities and reduce the segmentation and labelling errors.

## 6. REFERENCES

- [1] Nazim Haoachine et al., “Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery,” in *ISMAR*, Oct 2013, pp. 199–208.
- [2] Rosalie Plantefeve et al., “Automatic alignment of pre and intraoperative data using anatomical landmarks for augmented laparoscopic liver surgery,” in *Biomedical Simulation*, vol. 8789, pp. 58–66. 2014.
- [3] Toby Collins et al., “Computer-assisted laparoscopic myomectomy by augmenting the uterus with pre-operative mri data,” in *ISMAR*, Sept 2014, pp. 243–248.
- [4] Alexandre Krupa et al., “Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing,” *Robotics and Automation, IEEE Transactions on*, vol. 19, no. 5, pp. 842–853, Oct 2003.
- [5] M. Allan et al., “Toward detection and localization of instruments in minimally invasive surgery,” *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 4, pp. 1050–1058, April 2013.
- [6] Y. Shu et al., “Segmentation of laparoscopic images: Integrating graph-based segmentation and multistage region merging,” in *2nd Canadian Conference on Computer and Robot Vision*, 2005, pp. 429–436.
- [7] B.V. Dhandra et al., “Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation,” in *ICPR*, 2006, pp. 695–698.
- [8] A. Chhatkuli et al., “Live image parsing in uterine laparoscopy,” in *ISBI*, April 2014, pp. 1263–1266.
- [9] Kristina Prokopetc et al., “Automatic detection of the uterus and fallopian tube junctions in laparoscopic images,” in *IPMI*, vol. 9123, pp. 552–563. 2015.
- [10] Masoud Nosrati et al., “Efficient multi-organ segmentation in multi-view endoscopic videos using pre-operative priors,” in *MICCAI 2014*, vol. 8674, pp. 324–331. 2014.
- [11] Lena Maier-Hein et al., “Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery,” *Medical Image Analysis*, vol. 17, pp. 974–996, 2013.
- [12] David Levin, “Mesh-independent surface interpolation,” in *Geometric Modeling for Scientific Visualization*, pp. 37–49. 2004.